

Claims:

- 1 **1.** A method of training a scoring matrix for use by a classification system, the
2 classification system for use in performing classification requests based on natural
3 language text and with use of said scoring matrix which has been based on a set of
4 training data comprising natural language text, the method comprising the steps of:
5 generating an initial scoring matrix comprising a numerical value for each of a
6 set of n classes in association with each of a set of m features, the initial scoring matrix
7 based on said set of training data and, for each element of said set of training data,
8 based on a subset of said features which are comprised in said natural language text of
9 said element of said set of training data and on one of said classes which has been
10 identified therefor; and
11 based on the initial scoring matrix and said set of training data, generating a
12 discriminatively trained scoring matrix for use by said classification system by
13 adjusting one or more of said numerical values such that a greater degree of
14 discrimination exists between competing ones of said classes when said classification
15 requests are performed, thereby resulting in a reduced classification error rate.
- 1 **2.** The method of claim 1 wherein said step of adjusting said numerical values is
2 performed with use of a Generalized Probabilistic Descent algorithm.
- 1 **3.** The method of claim 2 wherein said step of adjusting said numerical values
2 comprises iteratively adjusting said numerical values until a stopping criterion is met.
- 1 **4.** The method of claim 3 wherein said stopping criterion comprises an empirical
2 loss threshold .
- 1 **5.** The method of claim 3 wherein said stopping criterion comprises a classification
2 error rate threshold.

1 **6.** The method of claim 2 wherein said step of adjusting said numerical values
 2 comprises, for each element of said set of training data, modifying values associated
 3 with the identified class and values associated with one or more of the other classes
 4 such that a score obtained for said element of said set of training data based on said
 5 modified values associated with the identified class is improved relative to one or more
 6 scores obtained for said element of said set of training data based on said modified
 7 values associated with said one or more other classes.

1 **7.** The method of claim 6 wherein said values are modified based on the following
 2 equation of the form:

$$r_{vw}(t+1) = \begin{cases} r_{vw}(t) + \varepsilon_i \frac{\partial l_k}{\partial d_k} x_w & \text{if } v = k \\ r_{vw}(t) - \frac{\varepsilon_i \frac{\partial l_k}{\partial d_k} x_w G_k(\bar{x}_i, R) (\bar{r}_v \cdot \bar{x}_i)^{\eta-1}}{\sum_{i \neq k} (\bar{r}_i \cdot \bar{x}_i)^\eta} & \text{if } v \neq k \end{cases}$$

1 **8.** The method of claim 1 wherein said classification requests comprise call routing
 2 requests, wherein said classes comprise call routing destinations, and wherein said set of
 3 training data comprises call routing requests together with associated identified call
 4 routing destinations therefor.

1 **9.** The method of claim 1 wherein said classification requests comprise document
 2 retrieval requests, wherein said classes comprise documents, and wherein said set of
 3 training data comprises document retrieval requests together with associated identified
 4 documents therefor.

1 **10.** The method of claim 1 wherein said natural language text upon which said
2 classification requests are based and said natural language text comprised in said set of
3 training data is processed without the use of stop word filtering.

1 **11.** A method of performing classification requests based on natural language text
2 and with use of a discriminatively trained scoring matrix which has been trained based
3 on a set of training data comprising natural language text, the scoring matrix having
4 been discriminatively trained by a method comprising the steps of:

5 generating an initial scoring matrix comprising a numerical value for each of a
6 set of n classes in association with each of a set of m features, the initial scoring matrix
7 based on said set of training data and, for each element of said set of training data,
8 based on a subset of said features which are comprised in said natural language text of
9 said element of said set of training data and on one of said classes which has been
10 identified therefor; and

11 based on the initial scoring matrix and said set of training data, generating said
12 discriminatively trained scoring matrix by adjusting one or more of said numerical
13 values such that a greater degree of discrimination exists between competing ones of
14 said classes when said classification requests are performed, thereby resulting in a
15 reduced classification error rate.

1 **12.** The method of claim 11 wherein said step of adjusting said numerical values is
2 performed with use of a Generalized Probabilistic Descent algorithm.

1 **13.** The method of claim 12 wherein said step of adjusting said numerical values
2 comprises iteratively adjusting said numerical values until a stopping criterion is met.

1 **14.** The method of claim 13 wherein said stopping criterion comprises an empirical
2 loss threshold .

1 **15.** The method of claim 13 wherein said stopping criterion comprises a
2 classification error rate threshold.

1 **16.** The method of claim 12 wherein said step of adjusting said numerical values
2 comprises, for each element of said set of training data, modifying values associated
3 with the identified class and values associated with one or more of the other classes
4 such that a score obtained for said element of said set of training data based on said
5 modified values associated with the identified class is improved relative to one or more
6 scores obtained for said element of said set of training data based on said modified
7 values associated with said one or more other classes.

1 **17.** The method of claim 16 wherein said values are modified based on the
2 following equation of the form:

$$r_{vw}(t+1) = \begin{cases} r_{vw}(t) + \varepsilon_l \frac{\partial l_k}{\partial d_k} x_w & \text{if } v = k \\ r_{vw}(t) - \frac{\varepsilon_l \frac{\partial l_k}{\partial d_k} x_w G_k(\bar{x}_l, R) (\bar{r}_v \cdot \bar{x}_l)^{\eta-1}}{\sum_{j \neq k} (\bar{r}_j \cdot \bar{x}_l)^\eta} & \text{if } v \neq k \end{cases}$$

1 **18.** The method of claim 11 wherein said classification requests comprise call
2 routing requests, wherein said classes comprise call routing destinations, and wherein
3 said set of training data comprises call routing requests together with associated
4 identified call routing destinations therefor.

1 **19.** The method of claim 11 wherein said classification requests comprise document
2 retrieval requests, wherein said classes comprise documents, and wherein said set of
3 training data comprises document retrieval requests together with associated identified
4 documents therefor.

1 **20.** The method of claim 11 wherein said natural language text upon which said
2 classification requests are based and said natural language text comprised in said set of
3 training data is processed without the use of stop word filtering.

1 **21.** An apparatus for training a scoring matrix for use by a classification system, the
2 classification system for use in performing classification requests based on natural
3 language text and with use of said scoring matrix which has been based on a set of
4 training data comprising natural language text, the apparatus comprising:

5 means for generating an initial scoring matrix comprising a numerical value for
6 each of a set of n classes in association with each of a set of m features, the initial
7 scoring matrix based on said set of training data and, for each element of said set of
8 training data, based on a subset of said features which are comprised in said natural
9 language text of said element of said set of training data and on one of said classes
10 which has been identified therefor; and

11 based on the initial scoring matrix and said set of training data, means for
12 generating a discriminatively trained scoring matrix for use by said classification
13 system by adjusting one or more of said numerical values such that a greater degree of
14 discrimination exists between competing ones of said classes when said classification
15 requests are performed, thereby resulting in a reduced classification error rate.

1 **22.** The apparatus of claim 21 wherein said means for adjusting said numerical
2 values executes a Generalized Probabilistic Descent algorithm.

1 **23.** The apparatus of claim 22 wherein said means for adjusting said numerical
2 values comprises means for iteratively adjusting said numerical values until a stopping
3 criterion is met.

1 **24.** The apparatus of claim 23 wherein said stopping criterion comprises an
2 empirical loss threshold .

1 **25.** The apparatus of claim 23 wherein said stopping criterion comprises a
2 classification error rate threshold.

1 **26.** The apparatus of claim 22 wherein said means for adjusting said numerical
2 values comprises, for each element of said set of training data, means for modifying
3 values associated with the identified class and values associated with one or more of the
4 other classes such that a score obtained for said element of said set of training data
5 based on said modified values associated with the identified class is improved relative
6 to one or more scores obtained for said element of said set of training data based on said
7 modified values associated with said one or more other classes.

1 **27.** The apparatus of claim 26 wherein said values are modified based on the
2 following equation of the form:

$$r_{vw}(t+1) = \begin{cases} r_{vw}(t) + \varepsilon_t \frac{\partial l_k}{\partial d_k} x_w & \text{if } v = k \\ r_{vw}(t) - \frac{\varepsilon_t \frac{\partial l_k}{\partial d_k} x_w G_k(\bar{x}_t, R) (\bar{r}_t \cdot \bar{x}_t)^{\eta-1}}{\sum_{j \neq k} (\bar{r}_t \cdot \bar{x}_t)^{\eta}} & \text{if } v \neq k \end{cases}$$

1 **28.** The apparatus of claim 21 wherein said classification requests comprise call
2 routing requests, wherein said classes comprise call routing destinations, and wherein
3 said set of training data comprises call routing requests together with associated
4 identified call routing destinations therefor.

1 **29.** - The apparatus of claim 21 wherein said classification requests comprise
2 document retrieval requests, wherein said classes comprise documents, and wherein
3 said set of training data comprises document retrieval requests together with associated
4 identified documents therefor.

1 **30.** The apparatus of claim 21 wherein said natural language text upon which said
2 classification requests are based and said natural language text comprised in said set of
3 training data is processed without the use of stop word filtering.

1 **31.** An apparatus for performing classification requests based on natural language
2 text and with use of a discriminatively trained scoring matrix which has been trained
3 based on a set of training data comprising natural language text, the scoring matrix
4 having been discriminatively trained by an apparatus comprising:

5 means for generating an initial scoring matrix comprising a numerical value for
6 each of a set of n classes in association with each of a set of m features, the initial
7 scoring matrix based on said set of training data and, for each element of said set of
8 training data, based on a subset of said features which are comprised in said natural
9 language text of said element of said set of training data and on one of said classes
10 which has been identified therefor; and

11 based on the initial scoring matrix and said set of training data, means for
12 generating said discriminatively trained scoring matrix by adjusting one or more of said
13 numerical values such that a greater degree of discrimination exists between competing
14 ones of said classes when said classification requests are performed, thereby resulting in
15 a reduced classification error rate.

1 **32.** The apparatus of claim 31 wherein said means for adjusting said numerical
2 values executes a Generalized Probabilistic Descent algorithm.

1 **33.** The apparatus of claim 32 wherein said means for adjusting said numerical
2 values comprises means for iteratively adjusting said numerical values until a stopping
3 criterion is met.

1 **34.** The apparatus of claim 33 wherein said stopping criterion comprises an
2 empirical loss threshold .

1 37. The apparatus of claim 16 wherein said values are modified based on the
2 following equation of the form:

1 **38.** The apparatus of claim 31 wherein said classification requests comprise call
2 routing requests, wherein said classes comprise call routing destinations, and wherein
3 said set of training data comprises call routing requests together with associated
4 identified call routing destinations therefor.

37

- 1 **40.** The apparatus of claim 31 wherein said natural language text upon which said
2 classification requests are based and said natural language text comprised in said set of
3 training data is processed without the use of stop word filtering.